5-2019

A Hidden Markov Factor Analysis Framework for Seizure Detection in Epilepsy Patients

Mahboubeh Madadi
University of Arkansas, Fayetteville

A Hidden Markov Factor Analysis Framework for Seizure Detection in Epilepsy Patients

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Statistics

by

Mahboubeh Madadi
University of Shahid Beheshti
Bachelor of Science in Applied Mathematics, 2005
University of Arkansas
Doctor of Philosophy in Industrial Engineering, 2015

May 2019
University of Arkansas

This thesis is approved for recommendation to the Graduate Council

---

Giovanni Petris, PhD
Thesis Director

Jyotishka Datta, PhD
Committee member

Avishek Chakraborty, PhD
Committee member

ABSTRACT

Approximately 1% of the world population suffers from epilepsy. Continuous long-term electroencephalographic (EEG) monitoring is the gold-standard for recording epileptic seizures and assisting in the diagnosis and treatment of patients with epilepsy. Detection of seizure from the recorded EEG is a laborious, time consuming and expensive task. In this study, we propose an automated seizure detection framework to assist electroencephalographers and physicins with identification of seizures in recorded EEG signals. In addition, an automated seizure detection algorithm can be used for treatment through automatic intervention during the seizure activity and on time triggering of the injection of a radiotracer to localize the seizure activity. In this study, we developed and tested a hidden Markov factor analysis (HMFA) framework for automated seizure detection based on different features such as total effective inflow which is calculated based on connectivity measures between different sites of the brain. The algorithm was tested on long-term (2.4-7.66 days) continuous sEEG recordings from three patients and a total of 16 seizures, producing a mean sensitivity of 96.3% across all seizures, a mean specificity of 3.47 false positives per hour, and a mean latency of 3.7 seconds form the actual seizure onset. The latency was negative for a few of the seizures which implies the proposed method detects the seizure prior to its onset. This is an indication that with some extension the proposed method is capable of seizure prediction.

ACKNOWLEDGEMENTS

I would like to thank Dr. Giovanni Petris for his endless support and patience and Dr. Avishek Chakraborty and Dr. Jyotishka Datta for serving as my committee members. Also, I would like to thank Dr. Leonidas Iasemidis for introducing me to this area of research, and providing me with the data. Last but not least, I would like to thank my husband, Shayan, and my parents for their continuous support and love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Overview

Epilepsy is a disorder characterized by spontaneous and recurrent seizures and afflicts nearly 1% (more than 50 million people) of the world's population [2], making it the fourth most common neurological disorder after Migraine, Stroke, and Alzheimer's [3]. The annual new cases are between 30 and 50 per 100,000 people in the general population [4]. Approximately in 30% of people with epilepsy, the condition is intractable to antiepileptic drugs [5].

Epileptic seizures are due to a sudden development of pathological, synchronous neuronal firing in the cerebrum and can be recorded by scalp, subdural and intracranial electrodes [6]. Seizures vary in duration (seconds to minutes), morphology and severity (clinical to subclinical, occurrence rate) within the same patient and across patients. Seizures may begin locally in portions of the cerebral hemispheres (partial/focal seizures) with a single or multiple foci, or simultaneously in both cerebral hemispheres (generalized seizures). After a seizures onset, partial seizures may remain localized and cause relatively mild cognitive, psychic, sensory, motor, or autonomic symptoms, or may spread (secondarily generalized) to cause altered consciousness, complex automatic behaviors, or bilateral tonic-clonic convulsions [6].

Previous studies have shown that seizures are not abrupt transitions in and out of an abnormal seizure state; instead, they follow a dynamical transition that evolves over minutes to hours [7, 8]. There is emerging evidence that the temporal dynamics of brain activity can be classified into four states: Interictal (between seizures, or baseline), Preictal (prior to seizure), Ictal (seizure), and Post-ictal (after seizures). The seizures (ictal states) cause temporary disturbances of brain functions (e.g., motor control, responsiveness, recall), for periods ranging from seconds to minutes. Seizures may be followed by a post-ictal period of

confusion or impaired sensorium that can last for several hours.

The task of seizure prediction and detection is an active area of research and can be achieved by using the Electroencephalography (EEG) signals to monitor the dynamical changes of the brain over time and intervene therapeutically at the right time. Seizure detection and especially prediction is difficult due to the presence of movement and other recording artifacts in EEG data. Difficulty in predicting seizures is one of the major factors affecting the quality of life of people with epilepsy. The primary challenge in seizure detection and prediction is differentiating between the ictal and non-ictal (pre-ictal, post-ictal, and inter-ictal) states from noise-contaminated data. Accurate seizure detection or prediction algorithms can provide warnings to encourage a patient to move to safety, or activate interventions such as electrical stimulation or drug delivery to avert seizures.

1.2 Electroencephalography (EEG)

Electroencephalography is the recording of the electrical activity produced by the firing of neurons within the brain and is the gold-standard for recording epileptic seizures and assisting in the diagnosis and treatment of patients with epilepsy. Electrical activities of the neurons are recorded through receptors, called electrodes, that are placed on the head and each one records the oscillations of brain electric potential on a specific part of the brain. The number of electrodes and positioning of them in the brain depends on the application. Typically, electrodes measure synaptic action averaged over tissue masses containing between roughly 100 million and 1 billion neurons. If the electrodes are placed on the surface of the scalp the EEG recording is called scalp or extracranial recording and has amplitude 10-100 μV. When measured on the surface or deep in the brain it has amplitude 10-20 mV and called intracranial EEG (iEEG).

EEG can be described in terms of rhythmic activity in specific frequency bands. The classification of EEG signals based on activity in specific frequency bands are discussed below:

Delta (δ) rhythm: EEG rhythmic activity below 4 Hz is categorized as delta rhythm. It is most prominent frontally in adults and posteriorly in children. It consists of high amplitude waves found during sleep and while performing tasks requiring continuous attention.

Theta (θ) rhythm: EEG activity in the frequency range 4-8 Hz is categorized as theta rhythm found in young children during sleep. This frequency range of EEG activity has been associated with reports of relaxed, meditative, and creative states.

Alpha (α) rhythm: EEG activity in the frequency range 8-13 Hz is categorized as alpha rhythm. It consists of regular waveforms with sharp peaks which are prominent in posterior regions of the head while resting. This was the first recorded electrical activity of the brain (recorded by Hans Berger); hence named as alpha rhythm.

Beta (β) rhythm: EEG activity in the frequency range 13-30 Hz is categorized as beta rhythm. It has symmetrical distribution on both sides of brain and is most evident frontally. Low amplitude beta with multiple and varying frequencies is often associated with active, busy or anxious thinking and active concentration.

Gamma (γ) rhythm: Gamma waves have the frequency range from 30 Hz and higher and are thought to represent binding of different populations of neurons together into a network for the purpose of carrying out a certain cognitive or motor function.

Figure 1.1 illustrates some examples of EEG signals of frequency bands Delta through Beta.

1.3 Conectivity Measures

Quantification of the interactions between different variables in a multicomponent system is critical for understanding the underlying dynamics of the system. Measures of connectivity in the frequency domain can provide robust estimates of the interactions be-

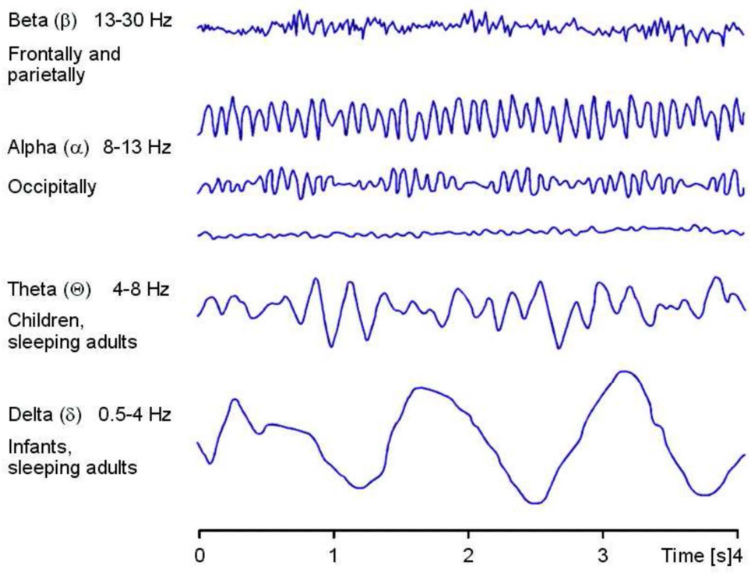

Figure 1.1: Examples of EEG signals of different frequency bands as indicated on the left (source http://www.bem.fi/book)

tween individual components of a system at specific frequencies. Connectivity measures, such as coherence, directed coherence (DC), directed transfer function (DTF), partial directed coherence (PDC), and generalized partial directed coherence (GPDC) have been applied to the study of brain dynamics. In the following the details of these connectivity measures are provided.

Coherence: Coherence is a statistic that can be used to examine the relation between two signals or time series [9]. Specifically, it estimates the extent to which two signals $x(t)$ and $y(t)$ may be predicted from one another by an optimum linear least squares function. The coherence between two signals $x(t)$ and $y(t)$ is a real-valued function that is defined as

$$C_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)}, \tag{1.1}$$

where $G_{xy}(f)$ is the cross–spectral density between x and y, and $G_{xx}(f)$ and $G_{yy}(f)$ are the auto–spectral density of x and y, respectively.

Directed Coherence (DC) and Directed Transfer Function (DTF): Let $\mathbf{X}(t) = \left[X_1(t), \ldots, X_n(t)\right]^T$ represent a set of simultaneously observed time series. Assume $\mathbf{X}(t)$ can be adequately represented by a vector Autoregressive Model of order p (VAR(p)); i.e.,

$$\mathbf{X}(t) = \sum_{\tau=1}^{p} \mathbf{A}(\tau)\mathbf{X}(t-\tau) + \epsilon(t) \tag{1.2}$$

where $\mathbf{A}(\tau)$ comprise the coefficients $a_{ij}(\tau)$ that relate the ij series at lag τ (describing the interactions between time series pairs over time) and $\epsilon(t) = \left[\epsilon_1(t) \ldots \epsilon_n(t)\right]^T$ is the vector of model innovations (with zero mean and covariance matrix Σ_ϵ).

Equation (1.2) can be transformed to describe relations in the frequency domain. After changing the sign of $\mathbf{A}$ and application of Z transform we have:

$$\begin{aligned} \epsilon(f) &= \mathbf{A}(f)\mathbf{X}(f) \\ \mathbf{X}(f) &= \mathbf{A}^{-1}(f)\epsilon(f) = \mathbf{H}(f)\epsilon(f) \\ \mathbf{H}(f) &= \left(\sum_{\tau=1}^{p} \mathbf{A}(\tau) e^{i-2\pi\tau f} \right)^{-1} \end{aligned} \tag{1.3}$$

The matrix of filter coefficients $\mathbf{H}(f)$ is called the transfer matrix of the system. Using the transfer matrix $\mathbf{H}(f)$, directed transfer function (DTF) [10] and directed coherence (DC) [11] are introduced as follows to describes the causal influence of channel j on channel k at frequency f:

$$\mathrm{DTF}_{j\to k}(f) = \frac{|H_{kj}(f)|^2}{\sqrt{\sum_{\ell=1}^{n} |H_{\ell j}(f)|^2}}. \tag{1.4}$$

$$\mathrm{DC}_{j\to k}(f) = \frac{\sigma_j |H_{kj}(f)|^2}{\sqrt{\sum_{\ell=1}^{n} \sigma_\ell^2 |H_{\ell j}(f)|^2}}. \tag{1.5}$$

where σ_k^2 denotes the variances of the innovations processes $\epsilon(t)$. Specifically, these measures

give the proportion of inflow from channel j to channel k to all the inflows to channel k.

Partial Directed Coherence (PDC) and Generalized Partial Directed Coherence (GPDC): Baccala et al. [12] formulated Partial Directed Coherence (PDC) which is a linear frequency-domain quantifier of the multivariate relationship between simultaneously observed time series for application in functional connectivity inference in neuroscience. Taking the discrete Fourier transform of both sides of (1.2), we obtain

$$\left(\mathbf{I} - \sum_{k=1}^{p} \mathbf{A}(\tau) e^{-i2\pi f\tau}\right)\mathbf{X}(f) = \mathbf{B}(f)\mathbf{X}(f) = \epsilon(f) \tag{1.6}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix, $\mathbf{X}(f)$, and $\epsilon(f)$ are the original signal and innovation vector process in the frequency domain, respectively. PDC is then expressed as

$$\mathrm{PDC}_{j\to k}(f) = \frac{\mathbf{B}_{kj}(f)}{\sqrt{\sum_{\ell=1}^{n} \mathbf{B}_{\ell j}(f)\mathbf{B}^{*}_{\ell j}(f)}}. \tag{1.7}$$

Since PDC suffers from correlated noise structures (mean and variance) in the innovations processes involved and it is not scale-invariant, Baccala et al. [13] formulated the concept of Generalized Partial Directed Coherence (GPDC) which is scale-invariant and more computationally efficient than other frequency-based connectivity measures:

$$\mathrm{GPDC}_{j\to k}(f) = \frac{\frac{1}{\sigma_k}\mathbf{B}_{kj}(f)}{\sqrt{\sum_{\ell=1}^{n} \frac{1}{\sigma_\ell^2}\mathbf{B}_{\ell j}(f)\mathbf{B}^{*}_{\ell j}(f)}} \tag{1.8}$$

where σ_k^2 denotes the variances of the innovations processes $\epsilon(t)$.

In general, a zero value in the $(k, j)^{th}$ entry in the matrix $\mathbf{B}(f)$ or $\mathbf{H}(f)$ indicates no directed interaction (no causal relation) from j to k.

1.4 Literature Review

Brain activity balances similarity between channels associated with propagation of information and dissimilarity associated with high entropy/information content. To this end, several multichannel methods have been applied to study clinical and experimental seizure activity using a variety of metrics such as cross correlation, coherence, Granger causality, transfer functions, and several nonlinear equivalents. Some researchers have focused on seizure detection from a functional connectivity perspective by exploring the synchronization abnormalities among neuronal population during the seizure period [14, 15, 16, 17, 18, 19, 20]. In the following we summarize the studies in the literature that use such measure for seizure detection/prediction.

1.4.1 Seizure Detection

Kerr et al. [21] used singular value decomposition of coherence connectivity matrices among the electrod sites to determine the dominant characteristics of the SEEG during the normal, the pre-ictal, and the ictal states. Omidvarnia et al. [22] used Dual extended Kalman Filter (DEKF) to estimate time varying MVAR parameters to compute time varying PDC and showed that seizures are detectable in neonatal epileptic patients. Wang et al. [23] proposed an approach based on the partial directed coherence, as a measure reflecting the physiological changes of brain activity before and after seizure onsets, to detect the seizure intervals of epilepsy patients. They calculated outflow information related to certain EEG channel by summing up the intensity of information flow propagated to other EEG channels in order to reduce the feature dimensionality. They applied support vector machine (SVM) classifier to detect ictal periods of data. In another work, Wang et al. [24] combined the wavelet decomposition and the directed transfer function (DTF) to develop a patient-specific seizure detection algorithm. Rana et al. [25] propose a seizure detection and analysis scheme based on the phase-slope index (PSI) which is a metric identifying an increases in the spatio temporal interactions between channels. They form a global metric of interaction between

channels and compare this metric to a threshold to detect the presence of seizures.

1.4.2 Seizure Prediction

It has long been observed that the transition from the interictal state (far from seizures) to the ictal state (seizure) is not sudden and may be preceded from minutes to hours by clinical, metabolic or electrical changes. Mormann et al. [15] reported an alteration in the degree of synchronisation between interictal and preictal EEG time series by means of mean phase coherence. Specifically, they characterized the degree of synchronization between EEG signals and retrospectively analyzed changes over time using the mean phase coherence as a measure for phase synchronization and the maximum linear cross correlation as a measure for lag synchronization. Santaniellot et al. [16] developed a hidden Markov model to find an optimal control-based quickest detection (QD) strategy to estimate the transition times from non-ictal to ictal states. They used connectivity measures (specifically the cross-power spectral density) as features to minimizes a cost function of detection delay and false positive probability. Winterhalder et al. [18] used eigenspectra of spacedelay correlation and covariance matrices from EEG data at multiple delay scales as features and used SVM to classify the patient's preictal or interictal states. Chaing et al. [26] developed an on-line retraining method with simple post-processing scheme based on wavelet coherence which is a measure of synchronization of the phase between channels. Wavelet coherence measures are calculated at different frequency bands and used as features for SVM classifier. To enhance their method they developed a post-processing scheme helps to reduce false positives rate. Schelter et al. [27] presented a classification method to minimize the false alarms adopting circadian concepts. The authors used the mean phase coherence as a seizure predictor, which causes an alarm to be raised if it exceeds a certain threshold.

1.4.3 Hidden Markov Models as a Classifier in Seizure Dtection and Prediction

Hidden Markov models (HMMs) are used for modeling a sequence of observable states with assumed hidden states. In HMMs the sequences of observable states are assumed to

be caused by some hidden states. The hidden states have a probability of emission for each of the finite set of observable states, and also a probability for changing to any other hidden states. There are a few studies in the literature of seizure detection/prediction that use variations of HMM as a classifier. For example, Hafidz et al. [28] developed a HMM classifier with a three state system including (i) ictal, (ii) preictal, and (iii) interictal states. They used stationary Wavelet Transform (SWT) to extract features from EEG signals. Baldassano et al. [29] developed a Bayesian nonparametric Markov switching process to parse intracranial EEG (iEEG) data into distinct dynamic event states. They modeled event state as a multidimensional Gaussian distribution. By detecting event states highly specific for seizure onset zones, their proposed method can identify precise regions of iEEG data associated with the transition to seizure activity. Extracting wavelet features of brain EEG, Esmaeili et al. [30] proposed an HMM with a mixture of Gaussian observation model as an unsupervised learning procedure to predict seizures, where the seizure predictions are derived from the posterior distributions over the hidden states in the HMM. They used Variational Bayesian (VB) method instead of the Maximum Likelihood estimation to train the proposed HMM. Santaniellot et al. [16] developed a hidden Markov model to find an optimal control-based quickest detection (QD) strategy to estimate the transition times from non-ictal to ictal states.

2 Tools and Methods

Figure 2.1 illustrates the overview of the proposed seizure detection framework. For each subject, EEG signals are first pre-processed. Pre-processing step includes segmentation of the signals into non-overlapping epochs and application of different filters. In the next step, we select the most informative channels by comparing signals randomly selected from icta and inter-ictal periods. Different features are then calculated for all the epochs at different frequency bands as well as for the full frequency. The proposed seizure detection method is then applied on different combinations of features, channels, and frequency bands. The combinations with the best performance in the testing procedure are then selected. The details of these steps are discussed below.

2.1 Signal pre-processing

Pre-processing is an important procedure in biomedical signal processing and analysis for identifying and removing noise or unwanted signals (e.g., biological signals and movement artifacts) or selecting a sub-band of frequency signal. EEG signals are first segmented into non-overlapping windows of length Δ. Next, to eliminate the noise coming from the power line an IIR notch filter is applied. We then apply band-pass Butterworth filter for analysis of data in different frequency bands.

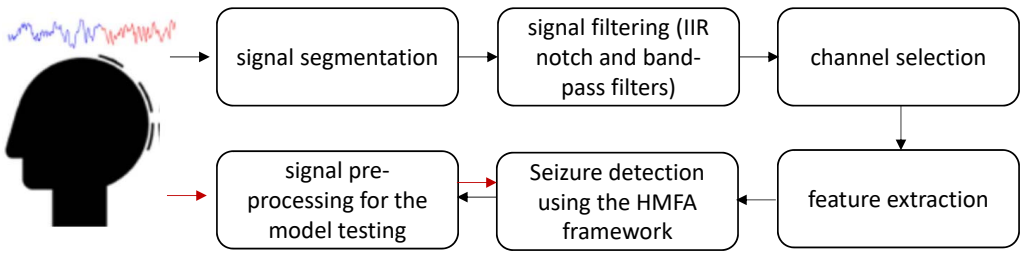

Figure 2.1: Overview of proposed seizure detection framework

2.2 Channel Selection

The aim of this step is to select EEG channels that provide the most discriminative information for identifying seizure onset in a specific patient. Channel selection reduces the computational time, eliminates irrelevant information, and improves the overall quality of the extracted features from EEG signals. The channel selection step includes two stages. In stage 1, for any pair of highly correlated channels, one of the correlated channels is randomly discarded. Removing correlated channels improves the estimation of the covariance matrix. In stage 2, the top p channels providing the most valuable information for detecting seizure are selected. More specifically, the difference in the variance of EEG signals in the ictal and inter-ictal states are calculated, and the top p channels that correspond to the highest difference of variance are selected.

2.3 Feature Extraction

We use different features including signal mean, variance, skewness, kurtosis, as well as a measure representing information flow between brain sites, namely effective inflow (EI) in the proposed seizure detection method.

Effective Inflow (EI):

Effective inflow is a measure of network connectivity proposed by Vlachos et al. [31] based on GPDC. Let $\bar{G}_{j\to i}(f_1, f_2)$ be the average GPDC over frequency band $[f_1 \quad f_2]$, i.e.,

$$\bar{G}_{j\to i}(f_1, f_2) = G_{j\to i}(f)_{f\in[f_1,f_2]} \tag{2.1}$$

Define the total inflow at node i as the sum of all 'flows' toward i from the rest of the nodes $j = 1, \ldots, n$, with $j \neq i$, as:

$$IG_i = \sum_{j=1, j\neq i}^{n} \bar{G}_{j\to i}(f_1, f_2). \tag{2.2}$$

Note that index IG_i in essence relates to the total effect exerted on i from all other sites.

In addition, we used first, second, third and fourth standardized moments, namely signal mean, variance, skewness, and kurtosis, as shown below, as features in the proposed seizure detection method. Let X_i^s be a segment of observation X_i, we have:

$$\bar{X_i^s} = \frac{1}{N}\sum_{h=1}^{N} X_{ih}^s, \quad i = 1, \dots n. \tag{2.3}$$

$$Var(X_i^s) = \frac{1}{N-1}\sum_{h=1}^{N} (X_{ih}^s - \bar{X_i^s})^2, \quad i = 1, \dots n. \tag{2.4}$$

$$\gamma(X_i^s) = \frac{\frac{1}{N-1}\sum_{h=1}^{N}(X_{ih}^s - \bar{X^s}_i)^3}{\left(\frac{1}{N-1}\sum_{h=1}^{N}(X_{ih}^s - \bar{X_i^s})^2\right)^3}, \quad i = 1, \dots n. \tag{2.5}$$

$$\kappa(X_i^s) = \frac{\frac{1}{N-1}\sum_{h=1}^{N}(X_{ih}^s - \bar{X_i^s})^4}{\left(\frac{1}{N-1}\sum_{h=1}^{N}(X_{ih}^s - \bar{X_i^s})^2\right)^2}, \quad i = 1, \dots n. \tag{2.6}$$

where N is the sample size in the data segment which is $N = \Delta \times sf$ with sf representing the signal sampling frequency.

2.4 Hidden-Markov Factor Analysis (HMFA)

In Hidden-Markov Factor Analysis (HMFA), a finite set of factor analyzers are used to model the relationship between the high-dimensional neural space and a low dimensional latent neural space, Factor analyzers at different time points are related with each other through a hidden Markov model (HMM). Please refer to Figure 2.2 [1].

Let $\mathbf{X}(t)$ denote the n-dimensional vector of the pre-processed electrode channels activity at time point $t \in \{1, \dots, T\}$, where n is the number of selected channels. Let $\mathbf{z}(t)$ denote the low-dimensional latent neural state of dimension $\ell \leq n$ and follows multivariate standard normal probability distribution, i.e.,

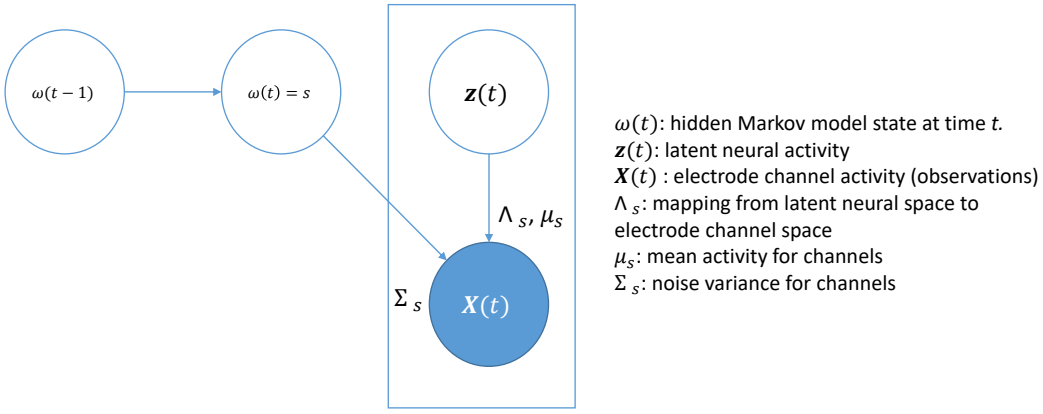

Figure 2.2: Hidden-Markov Factor Analysis [1]

$$\mathbf{z}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{2.7}$$

Also, assume $\omega(t) \in \{1, \ldots, S\}$ be hidden Markov model state at time t. We assume linear-Gaussian relationship between the electrode activity $\mathbf{X}(t)$ and the latent neural state $\mathbf{z}(t)$, that is

$$\mathbf{X}(t) \mid \mathbf{z}(t), \omega(t) = s \sim \mathcal{N}(\Lambda_s \mathbf{z}(t) + \mu_s, \Sigma_s) \tag{2.8}$$

where Λ_s is a $n \times \ell$ matrix and represents factor loading matrix, and μ_s, and Σ_s are the mean vector and covariance matrix. As it is standard in Factor Analysis (FA), we constrain the covariance matrix Σ_s to be diagonal, where the diagonal elements are the independent noise variances of each electrode.

The factor analyzers at different time points are related through an HMM to capture the idea of neural trajectories moving within as well as between low-dimensional subspaces. Start and transition probability parameters are learned for the HMM. The number of states in the HMM (S) must be determined a priori.

$$\pi_s = P(\omega(t) = s), \quad A_{jk} = P(\omega(t+1) = k | \omega(t) = j). \tag{2.9}$$

The parameters of the HMFA model $\Theta = \big(\pi, A, (\mu_s, \Sigma_s, \Lambda_s)_{s=1}^{S}\big)$ can be estimated using a variant of expectation maximization (EM) algorithm known as the Alternating Expectation Conditional Maximization (AECM) algorithm [1]. The algorithm seeks the model parameters that maximize the probability of the observations.

Assuming that there are m recordings for a specific patient (i.e., $i = 1, \ldots, m$) represented by $\mathbf{X}^i(t) = \big[X_1^i(t), \ldots, X_n^i(t)\big]^T$, the outline of the AECM training algorithm to obtain the model parameters, i.e., $\Theta = \big(\pi, A, (\mu_s, \Sigma_s, \Lambda_s)_{s=1}^{S}\big)$, is given below:

First E step: In this step, the factors $\mathbf{z}_i(t)$ are marginalized out and only the state labels remain as missing data in the expected log-likelihood:

$$\begin{aligned}\ell_1(\Theta) = &\sum_{i=1}^{m}\sum_{j=1}^{S} P\big(\omega_i(1) = j|\mathbf{X}^i\big)\log\pi_j \\ &+ \sum_{i=1}^{m}\sum_{t=1}^{T_i}\sum_{j=1}^{S}\sum_{k=1}^{S} P\big(\omega_i(t) = k|\omega_i(t-1) = j, \mathbf{X}^i\big)\log A_{jk} \\ &+ \sum_{i=1}^{m}\sum_{t=1}^{T_i}\sum_{j=1}^{S} P\big(\omega_i(t) = j|\mathbf{X}^i\big)\log P\big(\mathbf{X}^i(t)|\omega_i(t) = j\big)\end{aligned} \tag{2.10}$$

where $P\big(\omega_i(t) = j|\mathbf{X}^i\big)$ and $P\big(\omega_i(t) = k|\omega_i(t-1) = j, \mathbf{X}^i\big)$ can be calculated using the forward and backward algorithms for HMMs, and $P\big(\mathbf{X}^i(t)|\omega_i(t) = j\big)$ can be determined as follows:

$$\begin{aligned}&\log P\big(\mathbf{X}^i(t)|\omega_i(t) = j\big) \propto \\ &-\frac{1}{2}\log|\Lambda_j\Lambda_j' + \Sigma_j| - \frac{1}{2}\big(\mathbf{X}^i(t) - \mu_j\big)'(\Lambda_j\Lambda_j' + \Sigma_j)^{-1}\big(\mathbf{X}^i(t) - \mu_j\big)\end{aligned} \tag{2.11}$$

First CM-Step: In this step parameters subset $\Theta_1 = (\pi, A, \mu)$ can be estimated by maximizing $\ell_1(\Theta)$ with respect to these parameters, i.e.,

$$
\begin{aligned}
\pi_j &= \frac{\sum_{i=1}^{m} P\big(\omega_i(t)=j|\mathbf{X}^i\big)}{m} \\
A_{jk} &= \frac{\sum_{i=1}^{m}\sum_{t=2}^{T_i} P\big(\omega_i(t)=k|\omega_i(t-1)=j,\mathbf{X}^i\big)}{\sum_{i=1}^{m}\sum_{t=2}^{T_i}\sum_{k=1}^{S} P\big(\omega_i(t)=k|\omega_i(t-1)=j,\mathbf{X}^i\big)} \\
\mu_j &= \frac{\sum_{i=1}^{m}\sum_{t=1}^{T_i} P\big(\omega_i(t)=j|\mathbf{X}^i\big)\mathbf{X}^i(t)}{\sum_{i=1}^{m}\sum_{t=1}^{T_i} P\big(\omega_i(t)=j|\mathbf{X}^i\big)}
\end{aligned}
\tag{2.12}
$$

Second E-Step: In this step, both the factors and state labels are taken as missing data in the expected log-likelihood, i.e.,

$$
\begin{aligned}
\ell_2(\Theta) =& \sum_{i=1}^{m}\sum_{j=1}^{S} P\big(\omega_i(1)=j|\mathbf{X}^i\big)\log\pi_j \\
&+ \sum_{i=1}^{m}\sum_{t=1}^{T_i}\sum_{j=1}^{S}\sum_{k=1}^{S} P\big(\omega_i(t)=k|\omega_i(t-1)=j,\mathbf{X}^i\big)\log A_{jk} \\
&+ \sum_{i=1}^{m}\sum_{t=1}^{T_i}\sum_{j=1}^{S} P\big(\omega_i(t)=j|\mathbf{X}^i\big)\log P\big(\mathbf{X}^i(t)|\mathbf{z}^i(t),\omega_i(t)=j\big) \\
&+ \sum_{i=1}^{m}\sum_{t=1}^{T_i}\sum_{j=1}^{S} P\big(\omega_i(t)=j|\mathbf{X}^i\big)\log P\big(\mathbf{z}^i(t)\big)
\end{aligned}
\tag{2.13}
$$

Second CM-Step: Using the estimates from the first CM-Step, next we have the following updates for covariance and loading matrices, i.e.,

$$
\begin{aligned}
\Lambda_j =& \sum_{i=1}^{m}\sum_{t=1}^{T_i} P\big(\omega_i(t)=j|X^i\big)\big(\mathbf{X}^i(t)-\mu_j\big)\,\mathbb{E}\left[\mathbf{z^i(t)}|\mathbf{X^i(t)},\omega_i(t)=j\right] \\
& \cdot\left(\sum_{i=1}^{m}\sum_{t=1}^{T_i} P\big(\omega_i(t)=j|X^i\big)\cdot\mathbb{E}\left(\mathbf{z^i(t)z^i(t)}^{'}|\mathbf{X^i(t)},\omega_i(t)=j\right)\right)^{-1}, \\
\Sigma_j =& \frac{1}{\sum_{i=1}^{m}\sum_{t=1}^{T_i} P\big(\omega_i(t)=j|X^i\big)}\cdot\operatorname{diag}\Bigg[\sum_{i=1}^{m}\sum_{t=1}^{T_i} P\big(\omega_i(t)=j|X^i\big) \\
& \left(\big(\mathbf{X}^i(t)-\mu_j\big)\big(\mathbf{X}^i(t)-\mu_j\big)^{'}-\Lambda_j\,\mathbb{E}\left[\mathbf{z^i(t)}|\mathbf{X^i(t)},\omega_i(t)=j\right]\big(\mathbf{X}^i(t)-\mu_j\big)^{'}\right)\Bigg],
\end{aligned}
\tag{2.14}
$$

where

$$
\begin{aligned}
&\beta_j=\Lambda_j^{'}\left(\Sigma_j+\Lambda_j\Lambda_j^{'}\right)^{-1}, \\
&\mathbb{E}\left[\mathbf{z^i(t)}|\mathbf{X^i(t)},\omega_i(t)=j\right]=\beta_j(\big(\mathbf{X}^i(t)-\mu_j\big), \\
&\mathbb{E}\left[\mathbf{z^i(t)z^i(t)}^{'}|\mathbf{X^i(t)}\right]=\mathbf{I}-\beta_j\Lambda_j+\beta_j\big(\mathbf{X}^i(t)-\mu_j\big)\big(\mathbf{X}^i(t)-\mu_j\big)^{'}\beta_j^{'}.
\end{aligned}
\tag{2.15}
$$

2.5 Testing Procedure

Once the model parameters Θ are learned, we need to decode sequence of observations to sequence of hidden states (i.e., pre-ictal, post-ictal, and inter-ictal the optimal sequence of hidden states) to test our model. Specifically, given as input an HMFA with parameters Θ and a sequence of observations $\mathbf{X}=\big(\mathbf{X}(1),\ldots\mathbf{X}(T)\big)$, we desire to find the most probable sequence of states $\big(\omega(1),\omega(2),\ldots,\omega(T)\big)$. The most common decoding algorithms for HMMs is the Viterbi algorithm. Viterbi algorithm makes uses of a dynamic programming to find the most likely sequence of hidden states. The details of the Viterbi algorithm is given below.

Algorithm 1 Viterbi algorithm

procedure VITERBI($\mathbf{X}, \Theta$) create a path probability matrix `viterbi`$[S, T]$
    for `each state` s `from 1 to` S ▷ initialization step
        `viterbi`$[s, 1] \leftarrow \pi(s) * P(\mathbf{X}(1)|\omega = s)$
        `backpointer`$[s, 1] \leftarrow 0$ **do**
    end for
    for `each time step` t `from 1 to` T **do** ▷ recursion step
        for `each state` s `from 1 to` S **do**
            `viterbi`$[s, t] \leftarrow \max_{s'}$ `viterbi`$[s', t-1] * A_{s's} * P(\mathbf{X}(t)|\omega = s)$ `backpointer`[s,t] $\leftarrow \text{argmax}_{s'}$ `viterbi`$[s', t-1] * A_{s's}$
        end for
    end for
    `viterbi`$[s, T] \leftarrow \max_s$ `viterbi`$[s, t-1] * A_{s\omega(T)}$ ▷ termination
    `backpointer`$[s, T] \leftarrow \text{argmax}_s$ `viterbi`$[s', t-1] * A_{s\omega(T)}$
    return the backtrace path by following `backpointer` to states back in time from `backpointer` $[\omega(T), T]$
end procedure

3 Numerical Results

3.1 Data Source

Data analysis was performed using Stereoelectroencephalography (SEEG) data of three epilepsy patients collected at the University of Alabama (UAB), School of Medicine. SEEG is a method for invasive study of patients with refractory epilepsy. The study has approval from the institutional review board to perform analysis and publish de-identified data. The number of seizures in these three patients are 2, 5 and 9. The number of electrodes and channels vary across patients from 102 to 168. Patients were monitored in epilepsy monitoring units (EMU) for several days. The sampling frequency for patients 2 and 3 is at 2048 Hz, while a sampling frequency of 500 is used for the first patient. Characteristics of EEG data are shown in Table 3.1.

3.2 Data Analysis

A non-overlaping sliding window of length $\Delta = 10$ second of data (20480 or 5000 samples) are feature calculation. Prior to feature calculation, IIR notch filters with a rejection band of 60, 120, 180, and 360 Hz are applied to each sliding window to remove the noise coming from the power line. We then performed a 2nd order band-pass Butterworth filter. Specifically, we generated 0.1-4 Hz (δ band) signal, 4.1-8 Hz (θ band) signal, 8.1-12 Hz

Table 3.1: Characteristics of EEG data

Patient	**No. of seizures**	**No. of channels**	**Data length (hr)**	**Sampling rate**
1	2	102	183.82	500
2	9	168	96.35	2048
3	5	150	58.28	2048

(α band) signal, 12.1-30 Hz (β band) signal, 30-80 Hz (low γ band), and 80+ Hz (high γ band) signal. The complete spectrum (the whole frequency available for the maximum sampling rate) is also used. For each patient, we then randomly selected data from ictal and non-ictal period and randomly removed one channel from each pair of highly correlated channels. Comparing the variance of ictal and inter-ictal periods for each remaining channel, $p \in \{3, 4, 5\}$ channels with the highest difference in variance are then selected for feature calculations. Five different features, as discussed in Section 2.3, are derived for studying the characteristics and analyzing EEG signals. Features are computed for each sub-frequency band signals created from the original EEG data as well as for the full frequency.

The performance of the proposed HMFA framework on seizure detection framework is estimated using one-fold (i.e., leave-one-record-out) cross-validation scheme. More specifically, for each patient, one recording containing an ictal period is left aside for testing, and the seizure onset detection framework uses the $K-1$ remaining seizure records for training. Next, the sensitivity, latency, and false positives per hour are calculated using the seizure recording that was withheld from the training set. Sensitivity is the number of correctly detected seizure onsets over all the seizures occurrences. Latency represents the delay between seizure detection and seizure onset. False detection is the number of false alarms triggered during the testing process. This process is repeated K times until each of the seizure records has been tested once.

Training of HMFA model was initially performed setting the number of hidden states S equal to 4 to be consistent with the commonly used assumption of four states of the brain, namely inter-ictal, pre-ictal, ictal and post-ictal. The testing results, however, showed that using $S = 2$ and 3 yield superior results. This is a result of features showing similar behavior to ictal periods in mainly post-ictal and sometimes in pre-ictal periods. Note that the maximum dimension of observations is $n = p \times 5$ (which mean n varies between 15 and 25) since we considered five different features in our analysis. We also analyzed the impact of feature elimination since for some patients some of the features are not providing discriminative information. The number of latent factors was varied between 2 and $\ell = 0.5p \times 5$. All

the analyses were done under different frequency bands as well as the full frequency. The analyses are performed in Python 3.7.0 and Matlab 9.4.0.

3.3 Results

For some patients (namely patient 2) the total inflow feature shows a significant difference in the ictal and inter-ictal period. However, the difference in the ictal and post-ictal periods do not show a significant difference. Figure 3.1 shows the total effective inflow measures for patient 2 at full frequency well prior to and after the ictal period for seizures 1, 2, 3, 4, 7, and 8. However, the total effective inflow feature is not significantly different in the ictal and non-ictal period for all patients, namely patient 1. For such patients, other features are picked up by the algorithm for HMFA training. For patient 1, the EEG signal mean varies significantly in the ictal period when compared to the non-ictal period. Figure 3.2 shows channel 3 standardized mean feature for the two seizures in patient 1 about 30 minutes prior to after seizures at β frequency band.

The corresponding results for all three epileptic patients are given in Table 3.3. The reported results are the average of all testing instances. The proposed seizure onset detection framework was able to identify on average 96.3% of the seizure occurrences correctly. In term of the false positives, on average, the proposed method triggers a false alarm 3.47 per hour, which is relatively high, compared to the reported false positive rate of existing seizure detection methods in the literature. However, note that this is the first study analyzing these set of patients. Previous detection methods are mainly applied on a dataset collected at Childrens Hospital Boston-Massachusetts Institute of Technology (CHB-MIT) which is inherently different from this dataset in terms of the patients and seizures characteristics. In addition, the high false positive rates could be related to artifacts. There is a high chance that some environmental or extraphysiologic artifacts reduced the performance of seizure detection framework. Regarding latency, the proposed seizure detector was able to detect the majority of the seizure onsets in 3.7 seconds of seizure onset. In some cases, the proposed method detects the seizure prior to its onset. This shows that with some extension

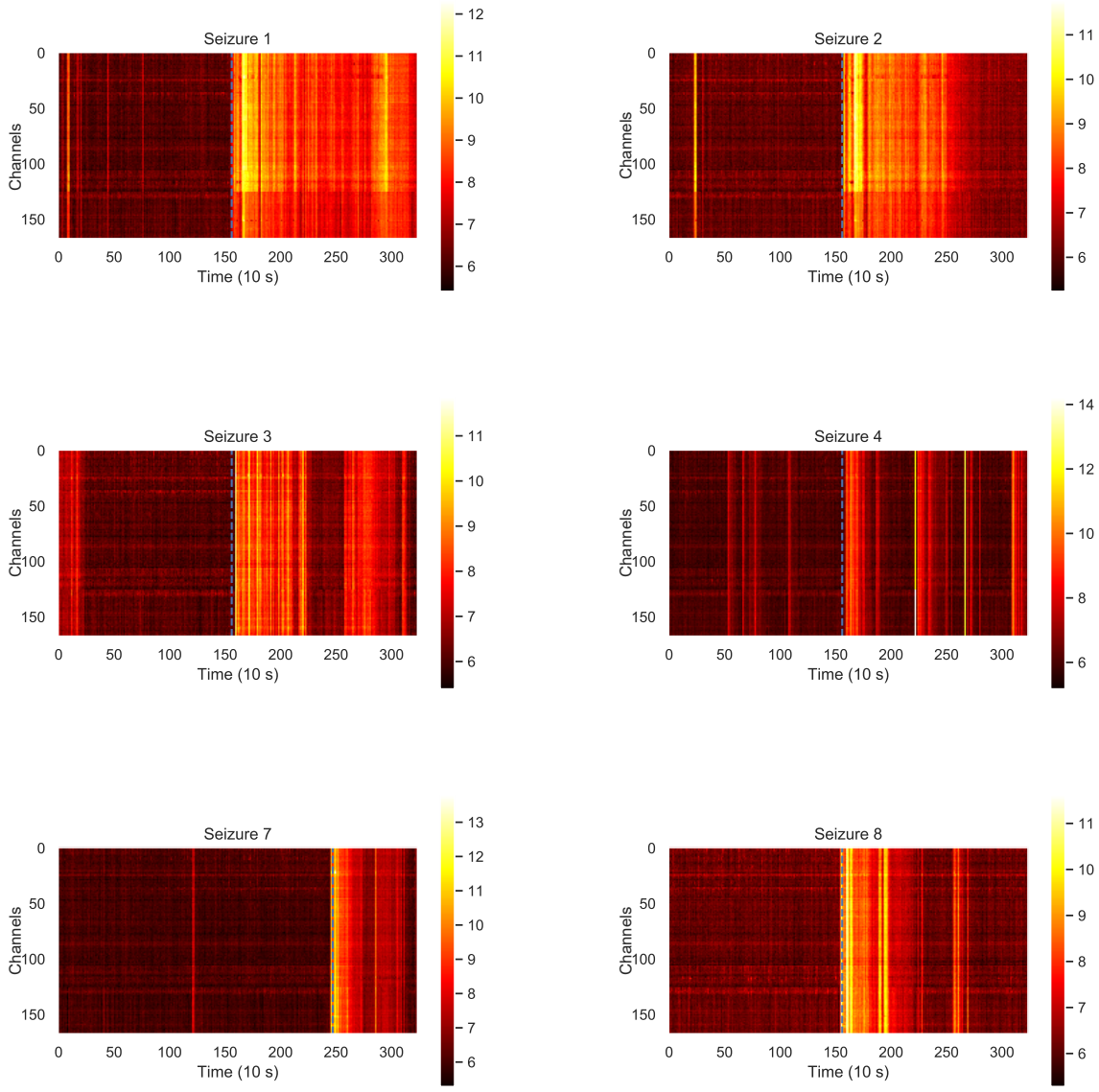

Figure 3.1: Total effective inflow measures for patient 2– the blue dotted lines represent the seizure onset

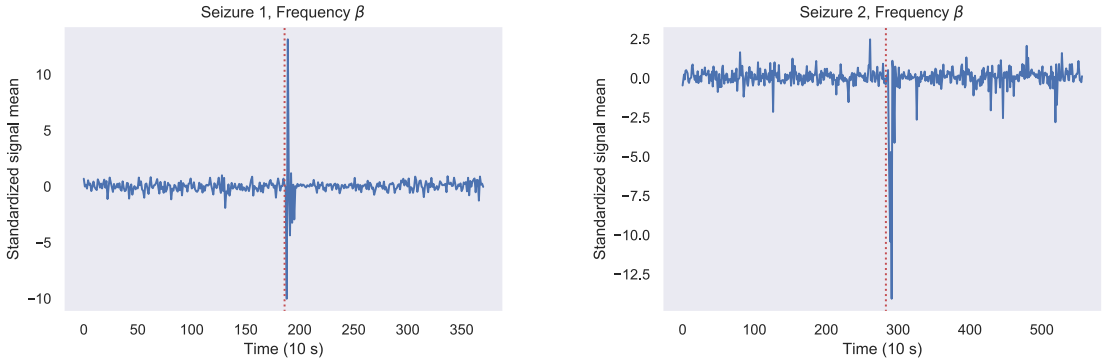

Figure 3.2: Standardized mean of channel 3 for patient 1 prior and after seizure. Red dotted lines represent the seizure onset

Table 3.2: Summary of the performance measures of the HMFA framework in seizure detection

Patient	**Sensitivity**	**False Positive (per hr)**	**Latency (seconds)**
1	2/2	2.6	10.1
2	8/9	3.3	7.5
3	5/5	4.1	-6.4

the proposed method is capable of seizure prediction. Figure 3.3 summarizes these results.

Figure 3.3: The performance of HMFA seizure detection framework on three patients from the UAB data

4 Conclusion

Epilepsy is a common neurological disorder characterized by abnormal excessive or synchronous neural activity in the brain. In this study, we develop an unsupervised learning method for seizure detection. Seizure onset detection enables developing implantable closed-loop devices for automatic intervention during the seizure activity and on time triggering of the injection of a radiotracer to localize the seizure activity, which could ultimately enhance the quality of life of epileptic patients. We developed and tested a hidden Markov factor analysis framework for seizure detection based on different measures. The primary measure of interest was total effective inflow to brain sites that are known or estimated to be the seizure onset zones. We also used other statistical measures such as signal mean, variance, skewness, and kurtosis as features in the HMFA seizure detection framework. The algorithm was tested on long-term (2.4-7.66 days) continuous SEEG recordings from three patients and a total of 16 seizures, producing a mean sensitivity of 96.3% across all seizures, a mean specificity of 3.47 false positives per hour, and a mean latency of 3.7 seconds form the actual seizure onset. The latency was negative for a few of the seizures which implies the proposed method detects the seizure prior to its onset. This is an indication that with some extension the proposed method is capable of seizure prediction. The proposed method is also capable of online seizure detection. The testing algorithm used in this thesis is the Viterbi algorithm. However, it is possible to implement a threshold based Bayesian updating framework which enables online seizure detection.

An extension of this algorithm can lead to a system for real-time detection and treatments. As for the future extension of this work, other features such as teager energy, short-term maximum Lyapunov exponent (ASTLmax) and entropy-based features could be used. More specifically, total inflow showed to be very informative for some patients (patient 2), but not for other patients (patient 1). Therefore using more informative patient-specific features,

the proposed seizure detection algorithm can be improved. In addition, as discussed above, the proposed model can be extended to an online threshed-based detection system through a Bayesian belief state updating. Further, incorporation of a control loop system optimized through partially observable Markov decision processes (POMDPs) framework could be used for treatment through production of electrical impulses that regulate abnormal impulses or injection of a radiotracer to localize the seizure activities.

Bibliography

[1] A. O. Omigbodun, "Finding structure in electrocorticographic neural signals for brain-machine interface applications," Ph.D. dissertation, UC San Diego, 2017.

[2] P. N. Banerjee, D. Filippi, and W. A. Hauser, "The descriptive epidemiology of epilepsya review," *Epilepsy research*, vol. 85, no. 1, pp. 31–45, 2009.

[3] D. Hirtz, D. Thurman, K. Gwinn-Hardy, M. Mohamed, A. Chaudhuri, and R. Zalutsky, "How common are the common neurologic disorders?" *Neurology*, vol. 68, no. 5, pp. 326–337, 2007.

[4] "World health organizations, epilepsy fact sheet," https://www.who.int/en/news-room/fact-sheets/detail/epilepsy, accessed: 2019-04-05.

[5] L. Ridsdale, J. Charlton, M. Ashworth, M. P. Richardson, and M. C. Gulliford, "Epilepsy mortality and risk factors for death in epilepsy: a population-based study," *Br J Gen Pract*, vol. 61, no. 586, pp. e271–e278, 2011.

[6] L. D. Iasemidis, "Epileptic seizure prediction and control," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 549–558, 2003.

[7] L. D. Iasemidis, D.-S. Shiau, J. C. Sackellares, P. M. Pardalos, and A. Prasad, "Dynamical resetting of the human brain at epileptic seizures: application of nonlinear dynamics and global optimization techniques," *IEEE transactions on biomedical engineering*, vol. 51, no. 3, pp. 493–506, 2004.

[8] J. C. Sackellares, L. D. IASEMIDIS, D.-S. Shiau, R. L. GILMORE, and S. N. ROPER, "Epilepsy–when chaos fails," in *Chaos in Brain?* World Scientific, 2000, pp. 112–133.

[9] S. L. Marple Jr and W. M. Carey, "Digital spectral analysis with applications," 1989.

[10] M. J. Kaminski and K. J. Blinowska, "A new method of the description of the information flow in the brain structures," *Biological cybernetics*, vol. 65, no. 3, pp. 203–210, 1991.

[11] Y. Saito and H. Harashima, "Tracking of information within multichannel eeg record-causal analysis in eeg," *Recent Advances in EEG and EMG Data Processing*, p. 133.

[12] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological cybernetics*, vol. 84, no. 6, pp. 463–474, 2001.

[13] L. A. Baccalá, K. Sameshima, and D. Takahashi, "Generalized partial directed coherence," in *2007 15th International conference on digital signal processing*. IEEE, 2007, pp. 163–166.

[14] L. D. Iasemidis, J. C. Sackellares, H. P. Zaveri, and W. J. Williams, "Phase space topography and the lyapunov exponent of electrocorticograms in partial seizures," *Brain topography*, vol. 2, no. 3, pp. 187–201, 1990.

[15] F. Mormann, R. G. Andrzejak, T. Kreuz, C. Rieke, P. David, C. E. Elger, and K. Lehnertz, "Automated detection of a preseizure state based on a decrease in synchronization in intracranial electroencephalogram recordings from epilepsy patients," *Physical Review E*, vol. 67, no. 2, p. 021912, 2003.

[16] S. Santaniello, S. P. Burns, A. J. Golby, J. M. Singer, W. S. Anderson, and S. V. Sarma, "Quickest detection of drug-resistant seizures: An optimal control approach," *Epilepsy & Behavior*, vol. 22, pp. S49–S60, 2011.

[17] R. B. Yaffe, P. Borger, P. Megevand, D. M. Groppe, M. A. Kramer, C. J. Chu, S. Santaniello, C. Meisel, A. D. Mehta, and S. V. Sarma, "Physiology of functional and effective networks in epilepsy," *Clinical Neurophysiology*, vol. 126, no. 2, pp. 227–236, 2015.

[18] M. Winterhalder, B. Schelter, T. Maiwald, A. Brandt, A. Schad, A. Schulze-Bonhage, and J. Timmer, "Spatio-temporal patient–individual assessment of synchronization changes for epileptic seizure prediction," *Clinical neurophysiology*, vol. 117, no. 11, pp. 2399–2413, 2006.

[19] I. Osorio and Y.-C. Lai, "A phase-synchronization and random-matrix based approach to multichannel time-series analysis with application to epilepsy," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 21, no. 3, p. 033108, 2011.

[20] M. Le Van Quyen, J. Soss, V. Navarro, R. Robertson, M. Chavez, M. Baulac, and J. Martinerie, "Preictal state identification by synchronization changes in long-term intracranial eeg recordings," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 559–568, 2005.

[21] M. S. Kerr, S. P. Burns, J. Gale, J. Gonzalez-Martinez, J. Bulacio, and S. V. Sarma, "Multivariate analysis of seeg signals during seizure," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 8279–8282.

[22] A. H. Omidvarnia, M. Mesbah, M. S. Khlif, J. M. O'Toole, P. B. Colditz, and B. Boashash, "Kalman filter-based time-varying cortical connectivity analysis of newborn eeg," in *2011 Annual international conference of the IEEE engineering in medicine and biology society*. IEEE, 2011, pp. 1423–1426.

[23] G. Wang, Z. Sun, R. Tao, K. Li, G. Bao, and X. Yan, "Epileptic seizure detection based on partial directed coherence analysis," *IEEE journal of biomedical and health informatics*, vol. 20, no. 3, pp. 873–879, 2016.

[24] D. Wang, D. Ren, K. Li, Y. Feng, D. Ma, X. Yan, and G. Wang, "Epileptic seizure detection in long-term eeg recordings by using wavelet-based directed transfer function," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 11, pp. 2591–2599, 2018.

[25] P. Rana, J. Lipor, H. Lee, W. Van Drongelen, M. H. Kohrman, and B. Van Veen, "Seizure detection using the phase-slope index and multichannel ecog," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 1125–1134, 2012.

[26] C.-Y. Chiang, N.-F. Chang, T.-C. Chen, H.-H. Chen, and L.-G. Chen, "Seizure prediction based on classification of eeg synchronization patterns with on-line retraining and post-processing scheme," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 7564–7569.

[27] B. Schelter, H. Feldwisch-Drentrup, M. Ihle, A. Schulze-Bonhage, and J. Timmer, "Seizure prediction in epilepsy: From circadian concepts via probabilistic forecasting to statistical evaluation," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 1624–1627.

[28] M. H. Abdullah, J. M. Abdullah, and M. Z. Abdullah, "Seizure detection by means of hidden markov model and stationary wavelet transform of electroencephalograph signals," in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*. IEEE, 2012, pp. 62–65.

[29] S. Baldassano, D. Wulsin, H. Ung, T. Blevins, M.-G. Brown, E. Fox, and B. Litt, "A novel seizure detection algorithm informed by hidden Markov model event states," *Journal of neural engineering*, vol. 13, no. 3, p. 036011, 2016.

[30] S. Esmaeili, B. Araabi, H. Soltanian-Zadeh, and L. Schwabe, "Variational bayesian learning for gaussian mixture hmm in seizure prediction based on long term eeg of epileptic rats," in *2014 21th Iranian Conference on Biomedical Engineering (ICBME)*. IEEE, 2014, pp. 138–143.

[31] I. Vlachos, B. Krishnan, D. M. Treiman, K. Tsakalis, D. Kugiumtzis, and L. D. Iasemidis, "The concept of effective inflow: application to interictal localization of the epileptogenic focus from ieeg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2241–2252, 2017.